ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	109628
ToLID	**xcIllCoin1**
Species	Illex coindetii
Class	Cephalopoda
Order	Oegopsida

Genome Traits	Expected	Observed
Haploid size (bp)	3,772,075,182	4,013,282,895
Haploid Number	6 (source: ancestor)	46
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q42

The following metrics were automatically flagged as below EBP recommended standards or different from expected:

. Observed Haploid Number is different from Expected

. Kmer completeness value is less than 90 for collapsed
. BUSCO single copy value is less than 90% for collapsed
. Not 90% of assembly in chromosomes for collapsed

Curator notes

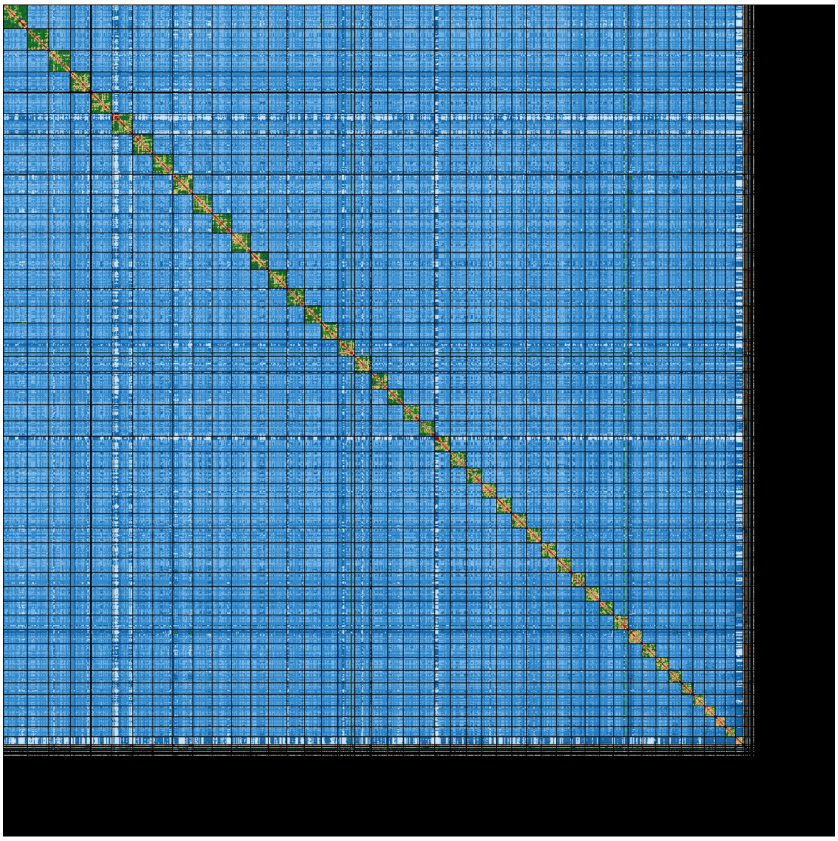
. Interventions/Gb: 11
. Contamination notes: ""
. Other observations: "The assembly of Illex coindetii (xcIllCoin1) is based on 40X PacBio data and Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr).The assembly process included the following steps: initial PacBio assembly generation with Hifiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. In total, 24 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 5.3 Mb (with the largest being 3.9Mb). Additionally, 1,395 regions totaling 286 Mb (with the largest being 2.7 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using ptgaul. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 4 haplotypic regions were removed,totaling 3.9 Mb (with the largest being 1.6 Mb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	4,017,277,276	4,013,282,895
GC %	33.93	33.93
Gaps/Gbp	181.47	184.14
Total gap bp	72,900	76,300
Scaffolds	717	706
Scaffold N50	74,066,118	75,295,916
Scaffold L50	23	23
Scaffold L90	63	56
Contigs	1,446	1,445
Contig N50	7,039,406	7,039,406
Contig L50	167	167
Contig L90	618	617
QV	42.5951	42.5966
Kmer compl.	82.3869	82.3515
BUSCO sing.	85.7%	85.7%
BUSCO dupl.	1.0%	1.0%
BUSCO frag.	8.9%	8.9%
BUSCO miss.	4.5%	4.5%

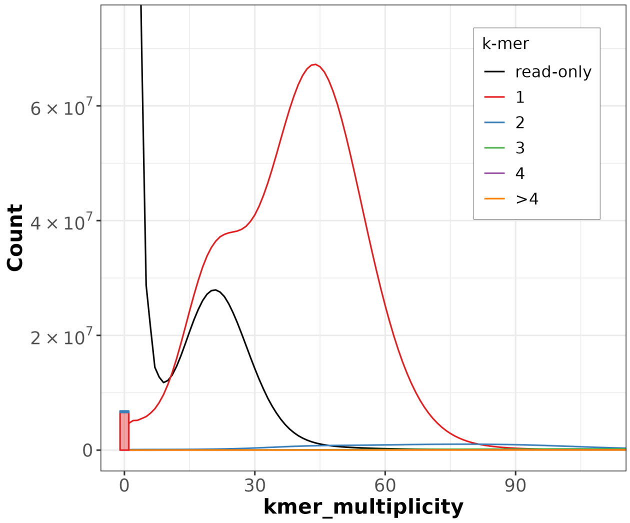
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

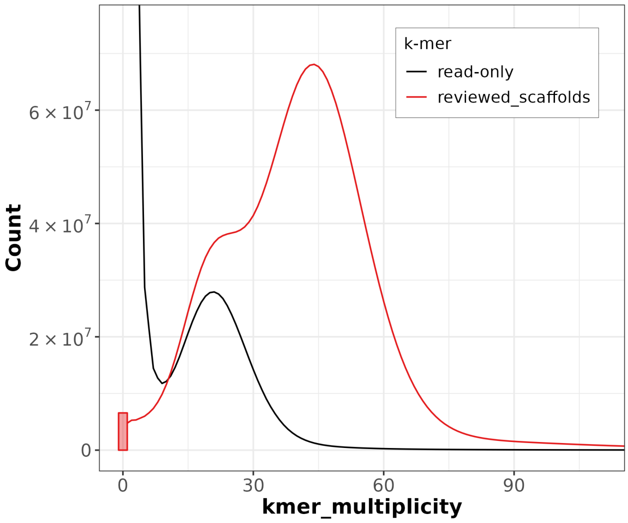


collapsed [LINK]

K-mer spectra of curated assembly

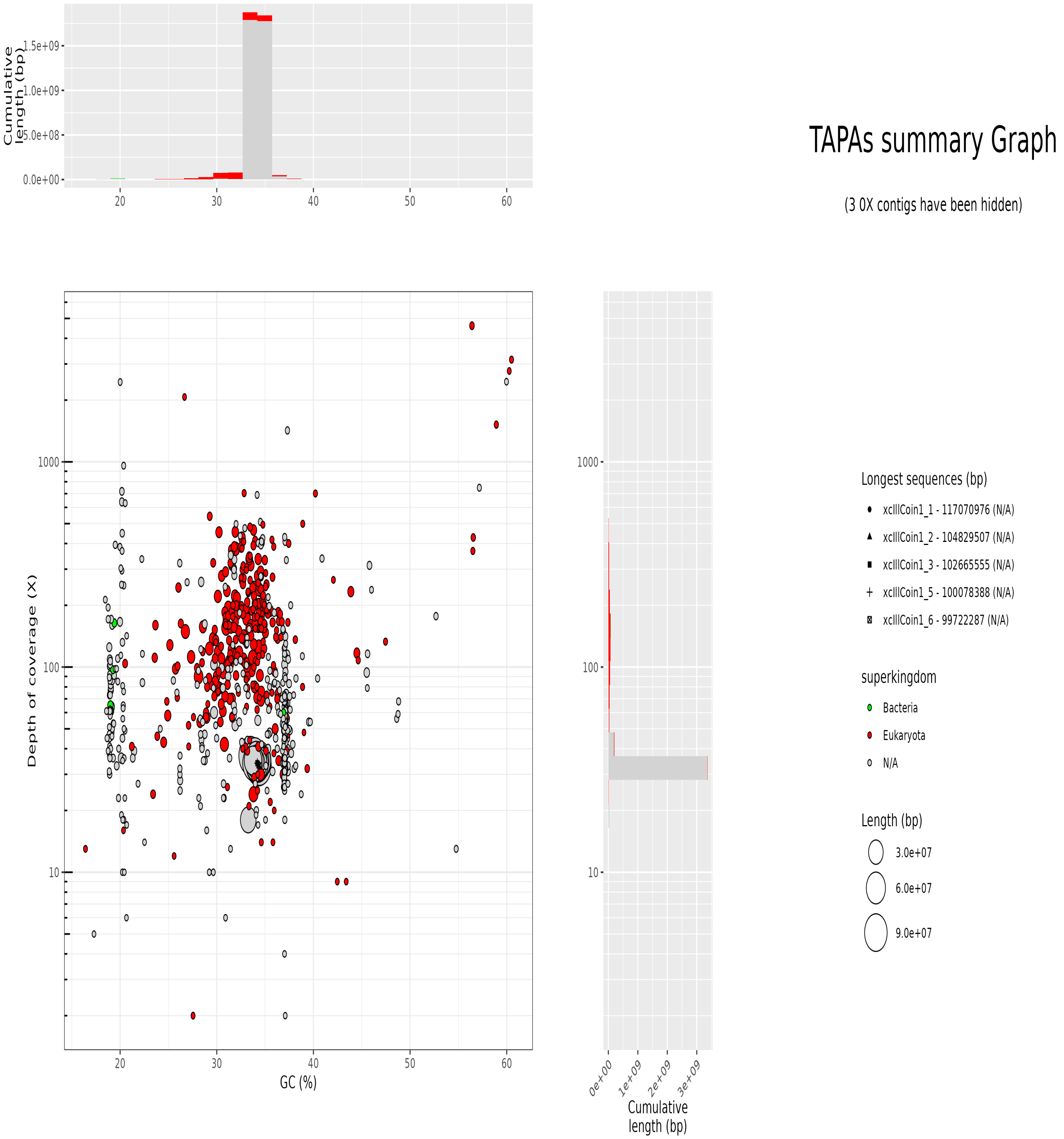


Distribution of k-mer counts per copy numbers found in asm



Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



collapsed. Bubble plot circles are scaled by sequence length, positioned by coverage and GC proportion, and coloured by taxonomy. Histograms show total assembly length distribution on each axis.

Data profile

Data	PACBIO Hifi	Arima
Coverage	40	165

Assembly pipeline

- **Hifiasm**
  |_ *ver:* 0.19.5-r593
  |_ *key param:* NA
- **purge_dups**
  |_ *ver:* 1.2.5
  |_ *key param:* NA
- **YaHS**
  |_ *ver:* 1.2
  |_ *key param:* NA

Curation pipeline

- **PretextMap**
  |_ *ver:* 0.1.9
  |_ *key param:* NA
- **PretextView**
  |_ *ver:* 0.2.5
  |_ *key param:* NA

Submitter: Caroline Menguy
Affiliation: Genoscope

Date and time: 2025-09-02 19:12:46 CEST